

APPENDIX 41

Date: May 27, 2010
To: Virginia Association of Municipal Wastewater Agencies
Maryland Association of Municipal Wastewater Agencies
From: Clifton F. Bell
Re: Recommendations on Baywide Loading Targets

On the May 24, 2010 Water Quality Goal Implementation Team (WQGIT) teleconference, USEPA presented the latest model results of dissolved oxygen (DO) and chlorophyll-*a* attainment under various loading scenarios. The USEPA announced its intention to derive the initial Baywide cap for nitrogen and phosphorous in the next week. At the conclusion of the teleconference, USEPA asked the states to provide quick feedback (1-2 days) via email on the appropriate Baywide target for main stem DO standards attainment.

The purpose of this memo is to address the present Baywide load allocation question. Highlights of this technical review are as follows:

- Given the high cost of management actions, it would be recommended to adopt an allocation approach that recognizes the proper uses and limitations of the modeling framework, and thus avoids:
 - Large swings in allocations between model versions.
 - Large swings in allocations to achieve numerically insignificant increases in attainment rates.
- The Bay Program modeling results should not be used in a manner that overestimates the precision of the model. The load-response predictions should be examined for asymptotic relations that would cause the target loads to be highly sensitive to small changes in nonattainment that exceed the precision of the model. In these cases, a difference in DO percent non-attainment rate of 3-5% should be used as a general guide to establish which model scenarios are "essentially equivalent."
- Based on the guideline cited above, the present Baywide nutrient load target should be based on the Target Load Option A scenario (200 Mlb/yr TN, 15 Mlb/yr TP). This loading recommendation was previously presented to and approved by the PSC.
- Due to modeling problems in shallow open water segments, as well as in embayments, the present model should not be used to adjust allocations for smaller local segments. At this time, the Baywide allocation process should be limited to deep

water and deep channel DO on the larger mainstem Bay problem segments including CB4MH, CB5MH, MD5MH, and VA5MH (as previously done).

- The Baywide allocation should not be based on side/tributary segments such as CHSMH, MAGMH, EASMH, or segments of the Elizabeth River, which can experience sensitivity local controls, natural causes of nonattainment, or local model calibration/resolution issues. In many cases, allocating to these smaller problem segments could require local modeling refinements. A phased approach is recommended to effectively address remaining problem segments.
- Both the magnitude and location of loads must remain primary considerations in the derivation of basin wide target loads. A further reduction of in loads from the southern tributaries (i.e., the James and York) would not significantly influence mainstem Bay DO attainment.

These specific recommendations are discussed in more detail below:

1. The allocation approach should recognize the proper use and limitations of the modeling framework. The Chesapeake Bay Program's framework of linked models, while very sophisticated, is still only an approximation of the natural system. The models were originally intended to provide an approximation of the large-scale hypoxic volumes under various loading scenarios. Under the present TMDL process, the model output is now being interpreted at spatial and temporal scales that exceed its precision. The ability of the model to distinguish small differences in attainment rates between model scenarios should be questioned considering factors such as:

- Continued instability in predicted attainment rates with each new version of the model.
- The documented ability of small numbers of outliers in the observed data set to cause predictions of non-attainment.
- The lack of a validation documenting the reliability of the model to accurately predict response to large-scale loading reductions that are simulated in the scenarios.

The model framework as a whole is conservative due to various assumptions such as:

- All point sources discharging maximum loads at all times.
- Selection of many conservative BMP efficiencies (as previously commented on by V/MAMWA).
- Allocations based on very small regions of the Bay system, which are much lower than needed for the great majority of the system.

Model uncertainty will be addressed by an implicit margin of safety associated with the conservativeness of the model. But in the context of choosing a Baywide load

allocation, the primary question becomes the following: *What is the ability of the model (or lack thereof) to truly differentiate attainment rates between scenarios?*

Consider the following hypothetical illustration: If Scenarios A, B, and C produce non-attainment rates of 25%, 6%, and 4%, one might believe that Scenarios B and C offer significant improvements from Scenario A. However, in reality Scenarios B and C are themselves essentially equivalent in terms of their response. It would be improper to make large cuts in allocations—with huge cost implications—on the basis of such small numerical differences in predicted attainment rates between B and C given the true sensitivity of the model.

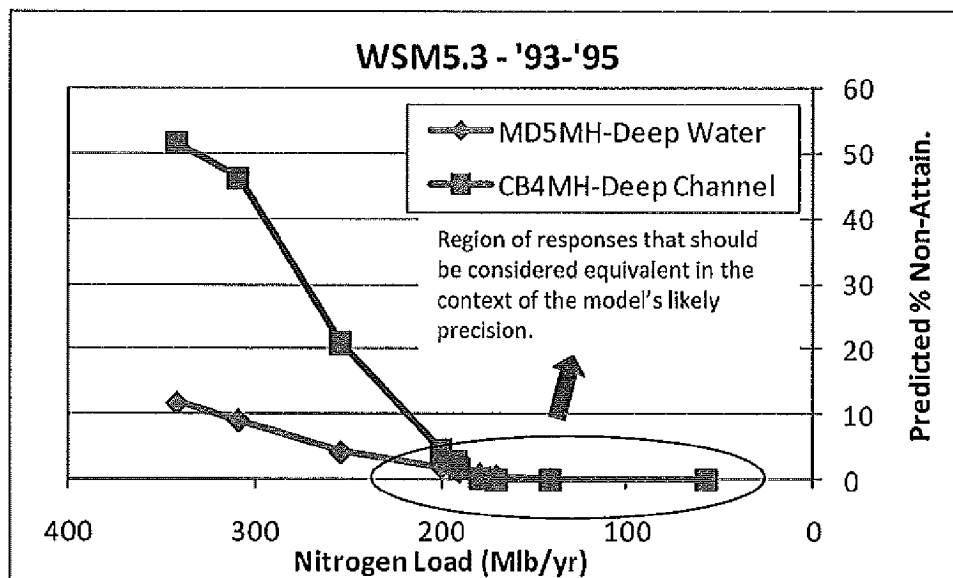


Figure 1: Example of asymptotic relations between predicted loading and non-attainment rates. Anywhere along the flat part of the curve, load allocations will be highly sensitive to very small changes in predicted non-attainment. These changes exceed the likely precision of the model.

2. The target load option A (200 Mlb/yr TN, 15 Mlb/yr TP) represents the load allocation at which key model segments are predicted to be in attainment with DO standards. Based on an examination of the most recent stoplight plots, the tributary strategy (200/15) appears to be the scenario at which key model segments are predicted to come into attainment, and/or the scenario beyond which more stringent scenarios are essentially equivalent in their water quality result (although vastly more stringent in terms of the associated management measures necessary to achieve that water quality result). The following is an examination of the key mainstem Bay segments (CB4MH and CB5MH) under the target load option A scenario:

- CB4MH
 - Deep water: In attainment
 - Deep channel: Within ~2% of attainment; equivalent to more stringent loading scenarios.
- CB5MH (Entire segment)
 - Deep water: In attainment
 - Deep channel: In attainment
- MD5MH (Maryland-only portion of CB5)
 - Deep water: Within ~2% of attainment; equivalent to more stringent loading scenarios.
 - Deep channel: In attainment
- VA5MH (Virginia-only portion of CB5)
 - Deep water: In attainment.
 - Deep channel: In attainment.

Considering the model limitations, it cannot be concluded that allocations lower than the target load option A would significantly improve attainment rates. Therefore, Target Load Option A is the most appropriate basis for the next Baywide loading target.

This loading scenario would not immediately address attainment of side segments including the CHSMH, MAGMH, and EASMH segments for deep water and deep channel DO. However, a comparison between results obtained by WSM5.1 and WSM5.3 indicate wide swings in response to attainment loading rates for these particular segments (Figure 2). Such wide swings indicate that further examination and explanation is needed to understand whether the model's predictions are scientifically defensible as a basis for decision making. Effectively addressing these segments might require separate, locally-oriented modeling analysis with a modeling tool better adapted to evaluating local conditions.

3. Due to open water modeling issues, the present model should not be used to adjust allocations for local segment open water DO attainment. The recent work by the Bay Program has highlighted serious model limitations in predicting attainment of open water DO standards. These include mechanistic errors in the simulation of DO in cells adjacent to shorelines, model grid resolution problems in small channels, bias in attainment rates due to a small number of unusual DO observations, and extrapolation of DO concentrations beyond the observed range.

Although mechanistic modeling problems are obvious in some segments due to unexpected load-response relationships, it should be stressed that the same mechanistic modeling problems are likely occurring in many other segments. Until and unless the model is shown to be accurate at simulating open water DO, and the level of accuracy is known, we recommend that these results not be used to further adjust target loads in a downward direction.

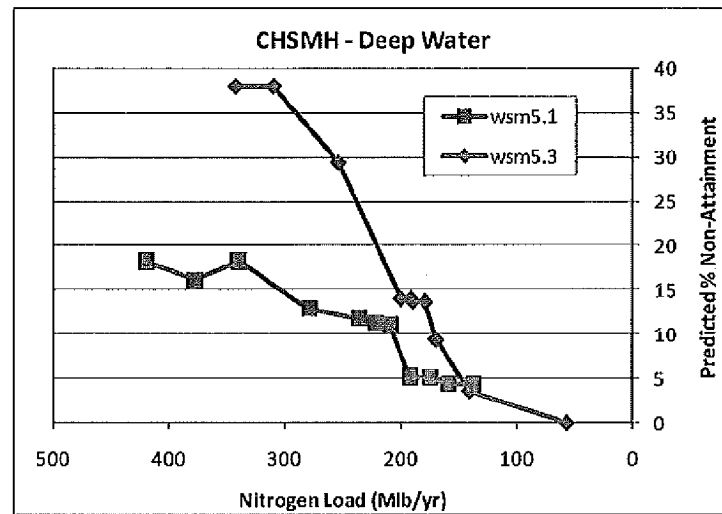


Figure 2: Example of widely different model predictions of non-attainment between model version 5.1 and 5.3 for a tributary segment.

4. *Both the magnitude and location of loads must remain primary considerations in the derivation of target loads.* The southern tributaries (i.e., the James and York) have very little effect on mainstem Bay attainment. Therefore, allocations in these basins should not be adjusted downwards to achieve a Baywide loading cap lower than 200/15. Although such adjustments might help achieve a given magnitude of loading, it would not achieve commensurate water quality benefits. Any adjustments to the magnitude of the Baywide loading cap should explicitly continue to consider the geography of load reductions.

5. *The Chesapeake Bay Program should achieve and communicate a clear understanding of the reasons for instability in predicted attainment rates between model version 5.1 and 5.3.* Based upon the premise that the water quality and sediment transport model (WQSTM) required little to no recalibration for use with watershed model (WSM) version 5.3, in comparison with WSM version 5.1, it is unclear why the different model versions would predict very different nonattainment rates at a given loading level for some segments (e.g., CHSMH, EASMH). The answer to this question is central to understanding whether the variation in predicted attainment rates is associated with manageable variables (e.g., the geography of load reductions) versus non-manageable variables (e.g., differences in the model algorithms). It would also help better quantify the amount of nonattainment that the model can truly distinguish between model scenarios. The Bay Program should diagnose the causes of the differences in model predictions, and clearly communicate these differences to the Bay partners before basinwide targets are selected.

cfb

Date: June 10, 2010
To: Virginia and Maryland Associations of Municipal
Wastewater Agencies
From: Clifton F. Bell
Re: Magnitude of Significant Differences in DO Criteria
Violation Rates

This technical memo presents the results of a power analysis to evaluate the minimum difference in dissolved oxygen (DO) violation rates that would be statistically detectable. Results indicated that differences in spatial violations rates of less than about 4-6% would not be measureable even over long monitoring periods. The conservative value might increase using other methods that consider intra-assessment period variability. Other parameters, such as chlorophyll-*a*, are much more temporally and spatially variable than DO, and thus could have significant higher MSD values.

BACKGROUND

The Chesapeake Bay Program's modeling framework is a tool to estimate what improvement in environmental conditions would result if certain nutrient management actions (i.e., scenarios) were put on the ground. Recent discussions with the Chesapeake Bay Program's (CBP) Water Quality Goal Implementation Team (WQGIT) have raised the question of whether the Chesapeake Bay modeling framework (and associated post-processing steps) can differentiate between model scenarios varying little in non-attainment rates. V/MAMWA raised this issue as part of a recommendation to avoid large swings in load allocations based model scenario predictions that are essentially equivalent given the likely precision of the model predictions (Bell, 2010).

The CBP's modeling framework is largely deterministic rather than stochastic, meaning that the predictions are based primarily on physical laws without explicit consideration of randomness or statistical variation. As a result, the actual precision of the non-attainment predictions for future management scenarios cannot be easily quantified. A related question that can be directly addressed is, "What is the minimum difference in non-attainment between two monitoring datasets that can be detected given data variability?" The question is relevant to the target load selection process because the public will have an expectation that the effect of the required controls on the environment be measurable and cost-effective, and water quality monitoring program data will be used to assess these improvements. Many stakeholders would agree that target loads should not be based on very small theoretical differences that exist between scenarios in the "model world" that would not translate to measureable improvements in the "real world".

Statistical power analysis represents a method to determine the magnitude of changes in water quality data that are statistically discernable from the variability that is inherent to the data. This technical memo presents one simple approach to determining the minimum significant differences in non-attainment rates of DO criteria.

POWER ANALYSIS

With a specific segment, month-to-month spatial violation rates are not normally distributed. However, the mean spatial violation rates for different 3-year assessment periods are approximately normally distributed, especially for segments that experience relatively low mean spatial violation rates (<10%). If the mean spatial violation rates for a segment were approximately normally distributed, the minimum significant difference (MSD) in the mean spatial violation rate could be determined by a parametric power analysis. This MSD of the spatial non-attainment rates would provide insight into the MSD of the area under the CFD assessment curve, because that area is calculated as a multiple of the spatial violation rates.

The MSD of spatial violations would be a direct function of the sample size (n), the Type I error rate (α), the Type II error rate (β), or power ($1 - \beta$), and standard deviation of the spatial violation rates. In this example, α and $1 - \beta$ were set to the conventional values of 0.05 and 0.8, respectively.

In the present example, the sample size n represents the number of three-year periods for which monitoring data are available before and after some treatment, such as the adoption of management practices. The MSD is inversely related to n , such that smaller differences in mean violation rates could be detected over longer monitoring periods. For the purposes of this exercise, n was set to 9, which corresponds to the number of independent (i.e., non-overlapping) three-year assessment periods over a twenty seven-year monitoring period. This approximately corresponds to the pre-TMDL period for which adequate monitoring data are available to assess spatial violation rates (early to mid 1980s – 2010). Hence, the power analysis will approximate the MSD that could be detected between pre-TMDL and post-TMDL monitoring periods up to about 2037.

In general, the standard deviation of spatial violation rates decreases as segments approach overall attainment relative to the CFD curve. For this example, it was desired to use a conservatively low standard deviation, to avoid overestimating the MSD. The selected values (3% and 4%) are typical of standard deviation of the mean spatial violation rates for deep water segments that are in overall attainment with the deep water CFD curve, as determined from a tabulation previously provided by the CBP (Attachment A).

The evaluation was conducted as a power analysis of a two-sample t-test using the software of Lenth (2010). Results indicate, under the assumption of this exercise, the MSD of the mean spatial violation rates is in the range of 4-6% (Table 1). In other words, under the assumptions specified for this analysis, the means of two independent

tabulations of mean spatial violation rates would have to differ by 4-6% before they could be determined to be statistically different, even over long monitoring periods. The actual difference in magnitude of overall (time-space) non-attainment rates would depend on the positions and shapes of the segment's curves relative to the respective reference curves. However, the differences in the mean percent area under their respective CFD curves would be highly correlated with the differences in spatial violation rates.

TABLE 1
Power Analysis: Two Sample t-test

Parameter	Value
Type I error rate (α)	0.05
Type II error rate (β)	0.8
Sample size (n); identical for both samples	9
Standard deviation in mean spatial violation rate	0.03 – 0.04
MSD	0.04 – 0.06

DISCUSSION

Based on the results of this analysis, segments that are in or close to attainment would have to have spatial DO violation rates that differ by 4-6% or more before they could be statistically distinguished from one another. Because the power analysis was conducted on the means of violations rates for three-year periods, the analysis did not consider variability of violation rates within three-year periods, which could increase the MSD. It would be recommended to explore other power analysis methods that consider intra-assessment period variation, such as analysis of variance (ANOVA) methods or methods that match observations by the month of measurement. However, the present analysis is analogous to the current assessment methodology by which a single non-attainment rate is estimated for each three-year period, without explicit consideration of the uncertainty of that value.

Results of the power analysis have indirect rather than direct bearing on the question of the precision of model non-attainment predictions. The model's precision is related to a host of factors other than the variability in the monitoring data, including the resolution of input datasets, calibration, variability associated with regressions developed for model post-processing, and variability between model versions. However, the power analysis does demonstrate that small (<4-6%) differences in model predictions of attainment between scenarios would probably not be measureable in the "real world".

This analysis is most pertinent to predictions of DO attainment in mainstem Bay segments. Other parameters, such as chlorophyll-*a*, are much more temporally and spatially variable than DO, and thus could have significant higher MSD values.

REFERENCES

- Bell, C.F. 2010. Recommendations on Baywide Loading Targets. Technical memorandum delivered to the Virginia and Maryland Associations of Wastewater Agencies. May 27, 2010. 5 p.
- Lenth, R. V. 2010. Java Applets for Power and Sample Size [Computer software]. Retrieved June 4, 2010, from <http://www.stat.uiowa.edu/~rlenth/Power>.

cfb

ATTACHMENT A

Spatial Violation Rates of Attaining Deepwater Segments

[Data from elec. comm., Excel file entitled "DO_violation_rates.xls", provided by J. Keisman to J. Pletl on 14 May 2009]

Segment	CB6PH	CB6PH	CB6PH	CB6PH	CB7PH	CB7PH	CB7PH	CB7PH	CB7PH	CB7PH	YRKPH
Period	96-98	97-99	99-01	04-06	96-98	97-99	98-00	99-01	00-02	04-06	04-06
Ranked Monthly Spatial Violations Rates - Deep Water 30-Day Mean Criterion	40%	51%	51%	40%	40%	40%	40%	26%	3%	23%	39%
	20%	40%	19%	27%	9%	26%	26%	3%	3%	21%	39%
	16%	20%	7%	25%	5%	9%	5%	3%	1%	5%	25%
	11%	9%	2%	20%	4%	5%	3%	0%	0%	3%	11%
	9%	2%	2%	9%	2%	1%	3%	0%	0%	1%	4%
	8%	0%	0%	3%	1%	1%	1%	0%	0%	1%	0%
	2%	0%	0%	3%	1%	0%	0%	0%	0%	0%	0%
	0%	0%	0%	3%	1%	0%	0%	0%	0%	0%	0%
	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%
	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
3 YR MEAN	9%	10%	7%	11%	5%	7%	7%	3%	1%	5%	10%
Note: 3 YR means are approximately normal with mean= 7% and standard deviation = 3.25% calculated using non-overlapping data											

Technical Rationale for Documenting Attainment for 1% Non-attainment Dissolved Oxygen Criteria Values

**State/District Co-Regulators
June 14, 2010 Conference Call
Attachment C2**

Rich Batiuk and Gary Shenk

Technical Rationale for 1%

- **Two separate analyses**
 - Evaluation of evidence for ‘residual’ of 1.5 percent or less non-attainment in dissolved oxygen criteria values across large ranges of model simulated load reductions across multiple Bay segments and designated uses (Batiuk)
 - Analysis of changes in the sensitivity of dissolved oxygen criteria attainment to simulated load reductions (Shenk)

Residual of 1.5% or Less Analysis

- Keep in mind:
 - ‘Stoplight plots’ already account for unallowable exceedences
 - Stoplight plots also account for restoration variances adopted in states’ WQS regulations
 - Stoplight plots are the result of a comprehensive analysis system: model scenario output -> regression generation -> transformation of monitoring data -> criteria assessment fully consistent with states/DC WQS regulations

Residual of 1.5% or Less Analysis

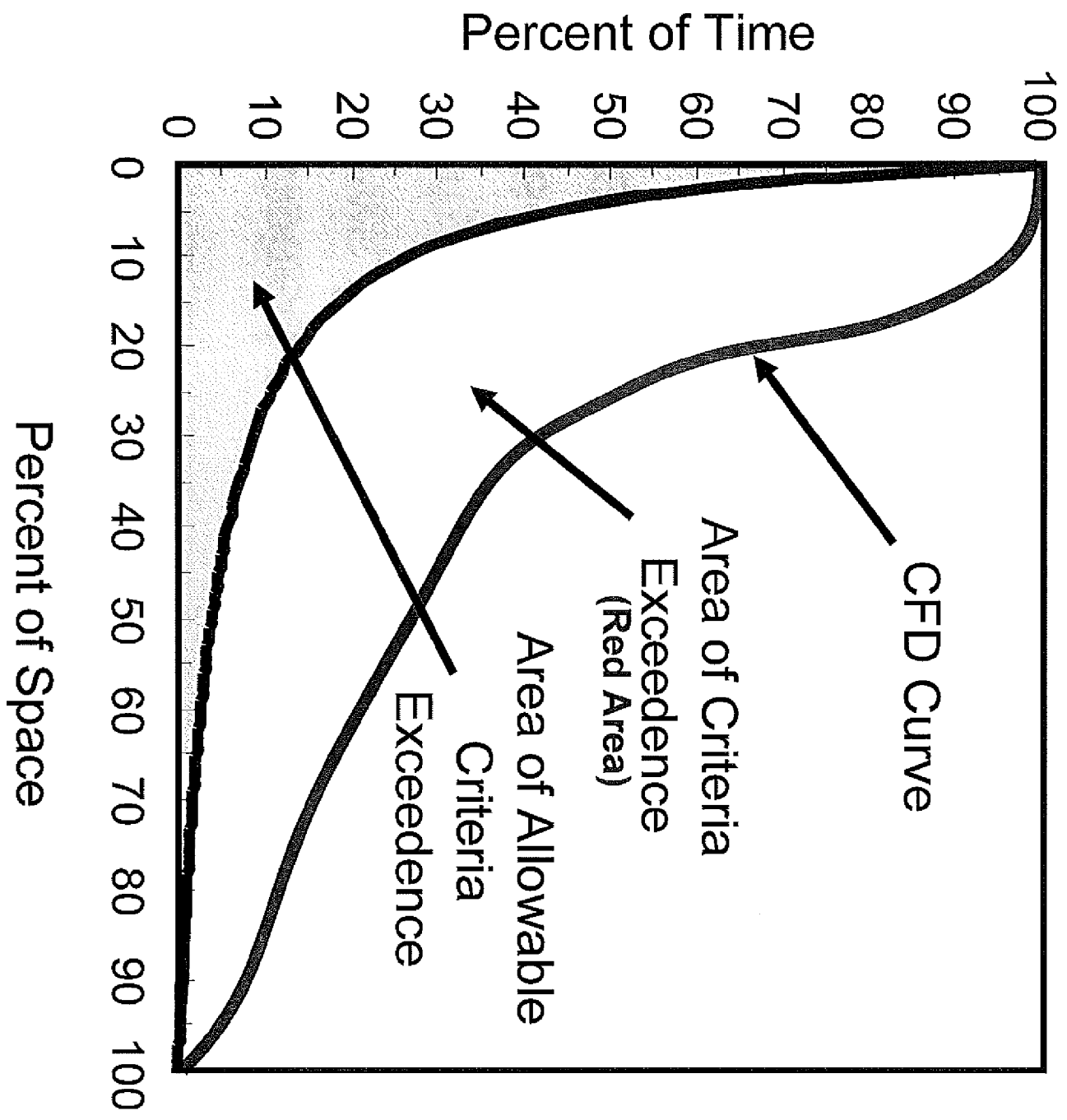
- 21 designated use-segments with non-attainment values ranging from 0.0% to 1.5% (will round down to 1%) based on May 24, 2010 stoplight plot presented to WQGIT
 - Mainstem Bay, major river, small tributaries and embayment segments all represented
 - 11 open-water, 8 deep-water and 2 deep-channel designated uses
- Model simulated nitrogen load reductions ranged from 9 to 151 million pounds

Chesapeake Bay Segment	Designated Use	Criteria Non-attainment Range (%)	Model Simulated Nitrogen Load Range (million pounds/yr)
CB7	Open-water	0.5-0.0	200-141
CHOMH1	Open-water	0.1-0.0	254-179
CSHMH	Open-water	0.8-0.1	342-309
DCA7F	Open-water	1.2-0.1	191-179
PAX7F	Open-water	1.0-0.6	190-179
DCPTF	Open-water	0.6-0.2	309-254
MAGMH	Open-water	1.3-0.3	342-191
MOBPH	Open-water	1.0-0.0	342-200
PIAMH	Open-water	0.1-0.1	191-179
TANMH	Open-water	1.5-0.1	342-309
YRKMH	Open-water	1.0-0.4	191-170
CB3MH	Deep-water	0.6-0.0	254-179
CB5MH	Deep-water	1.5-0.0	254-141
CHSMH	Deep-water	0.5-0.4	170-141
EASMH	Deep-water	0.8-0.2	200-170
MD5MH	Deep-water	1.5-0.1	191-141
MAGMH	Deep-water	0.5-0.5	170-141
PATMH	Deep-water	1.1-0.1	200-190
VAA5MH	Deep-water	0.7-0.0	254-179
CB3MH	Deep-channel	0.2-0.1	200-190
EASMH	Deep-channel	1.3-0.0	190-170

Sensitivity to Load Reductions

- Plotting the change in unallowable exceedence (red area) per loading unit against the starting red area
- Change in red area between two scenarios is divided by the change in the load
- Changes in N and P loads are combined into a single measure:

$$\text{Load units} = (N + 10 \cdot P) / 2$$



Sensitivity to Load Reductions

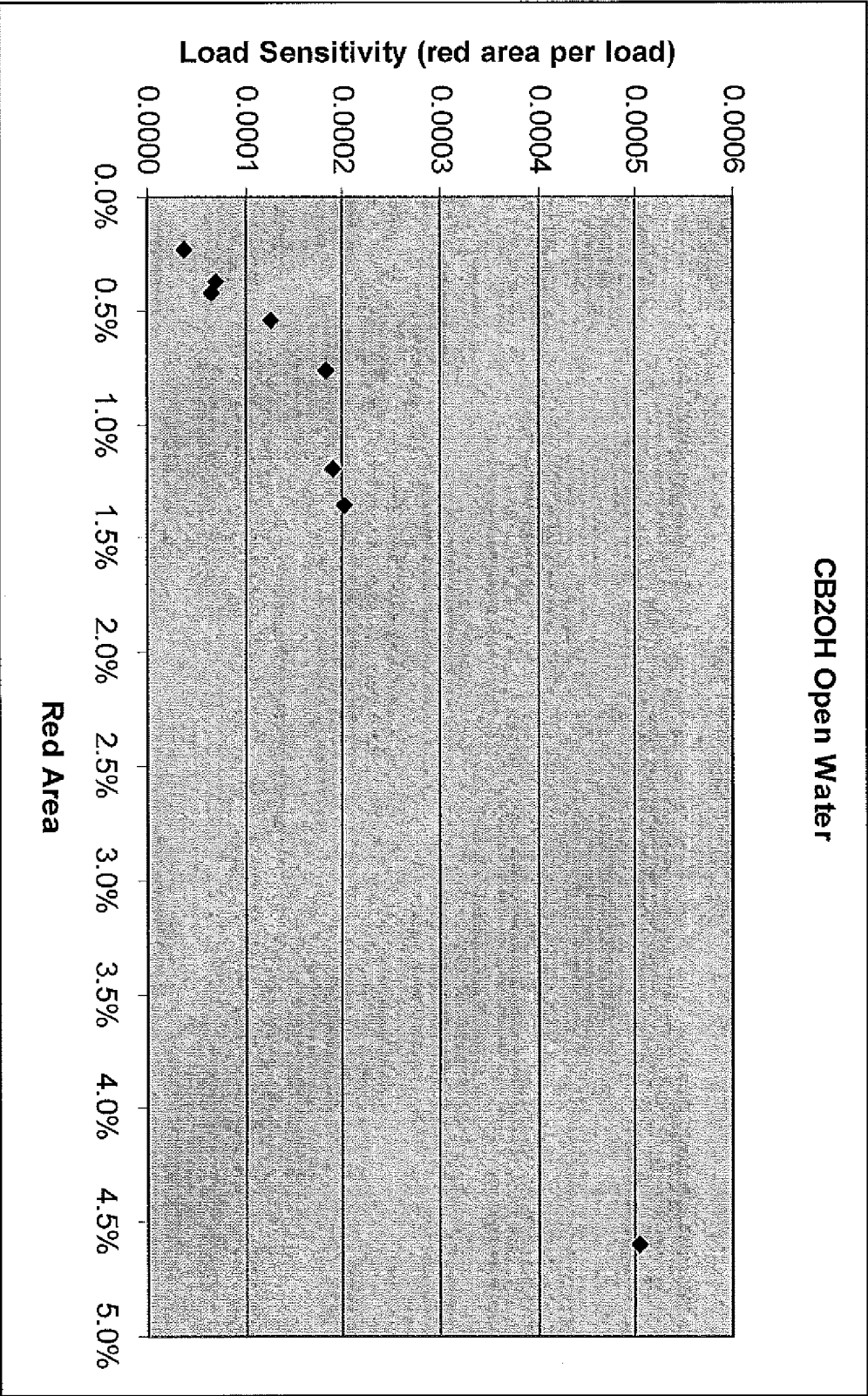
- When plotted against the starting red area, this allows for a direct comparison of sensitivity of the analysis system to load changes across different level of non-attainment

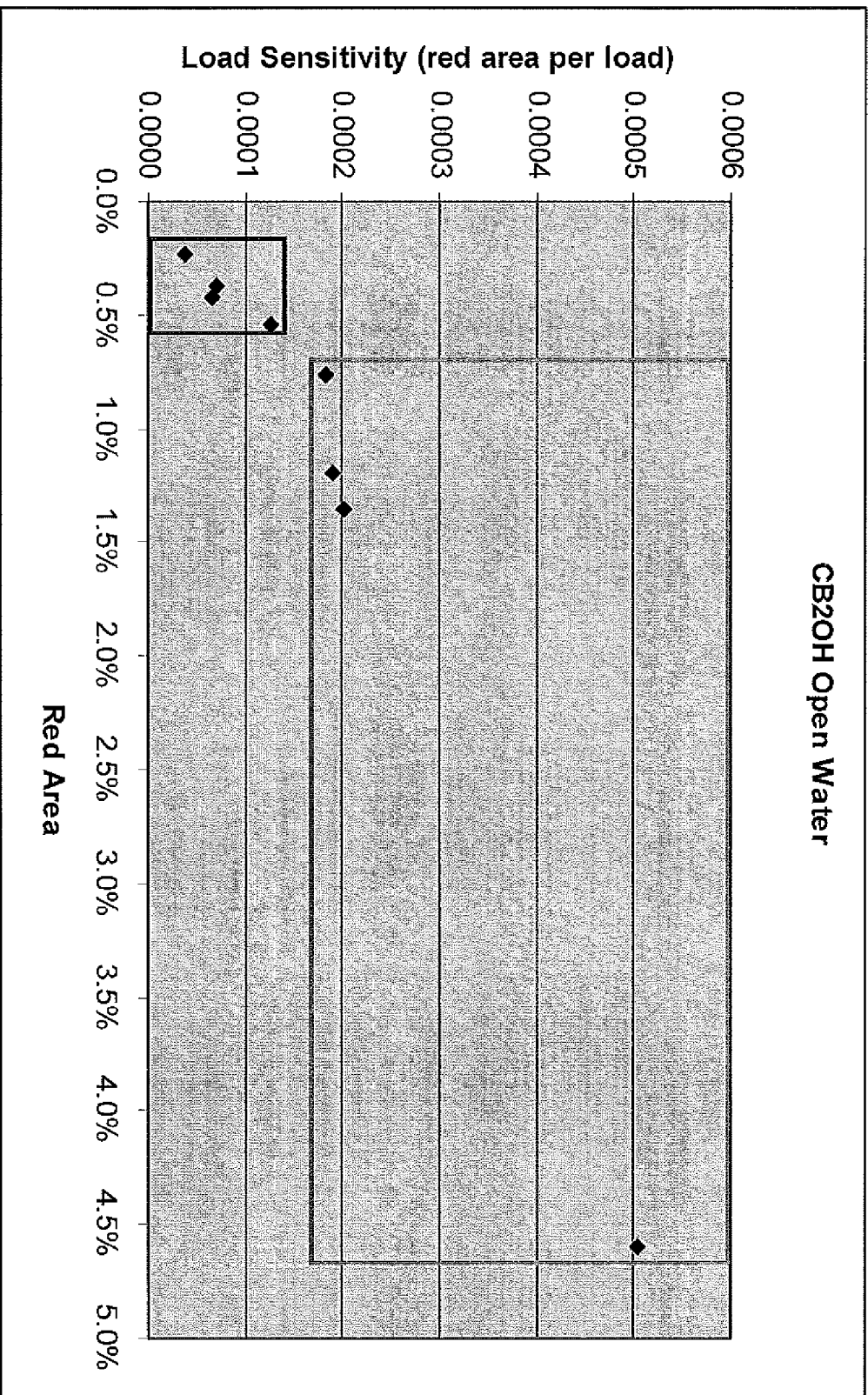
Sensitivity to Load Reductions

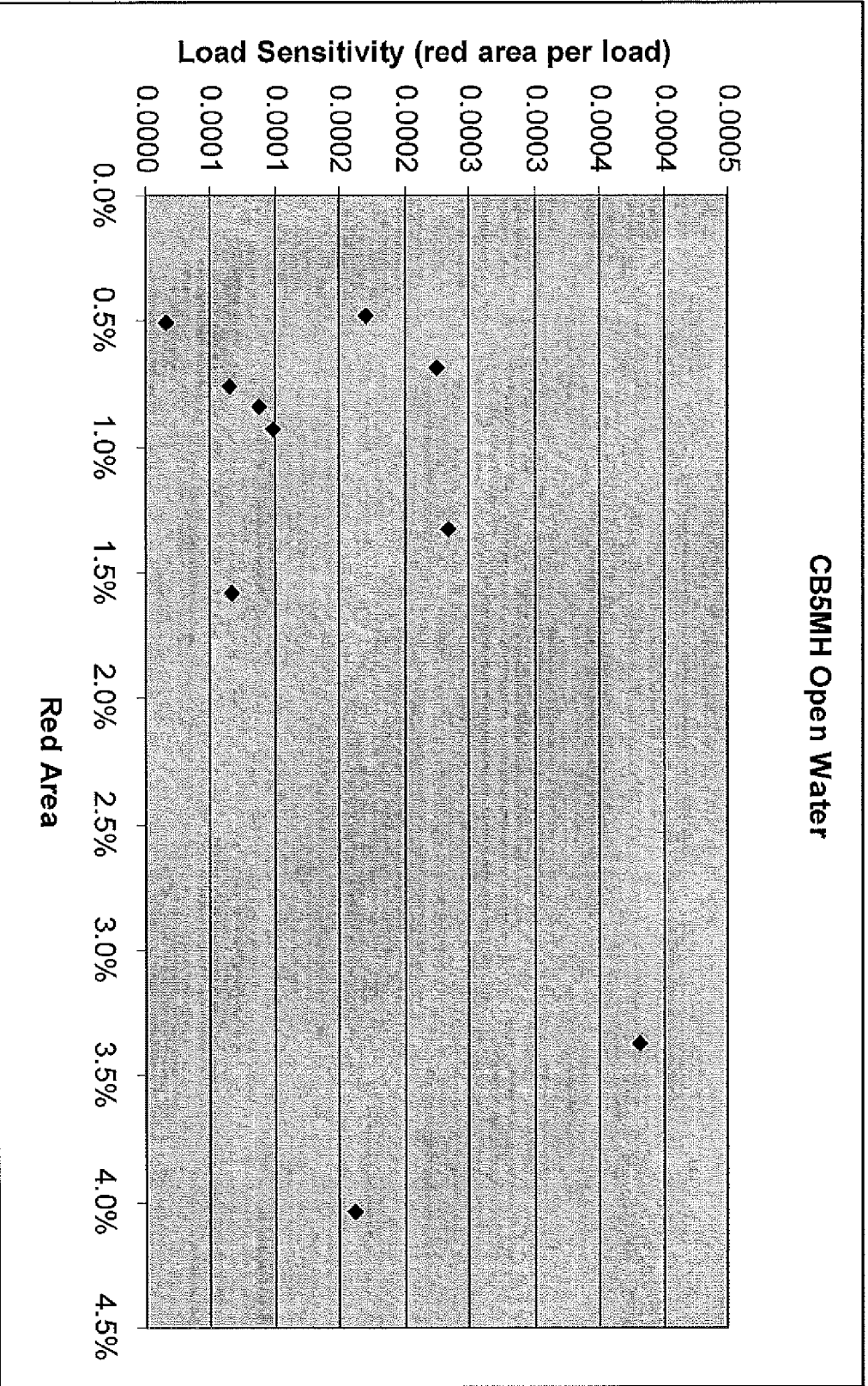
- 12 scenarios with eight 3-year periods for a total of 88 possible sensitivity assessments per designated use segment
 - Calculation involving scenarios where criteria were attained were not included in this analysis
- Not amenable to tidal tributary segments
 - Loading are baywide, not specific to tributaries
 - Existing scenarios used for analysis have varying levels of reduction between different tributaries

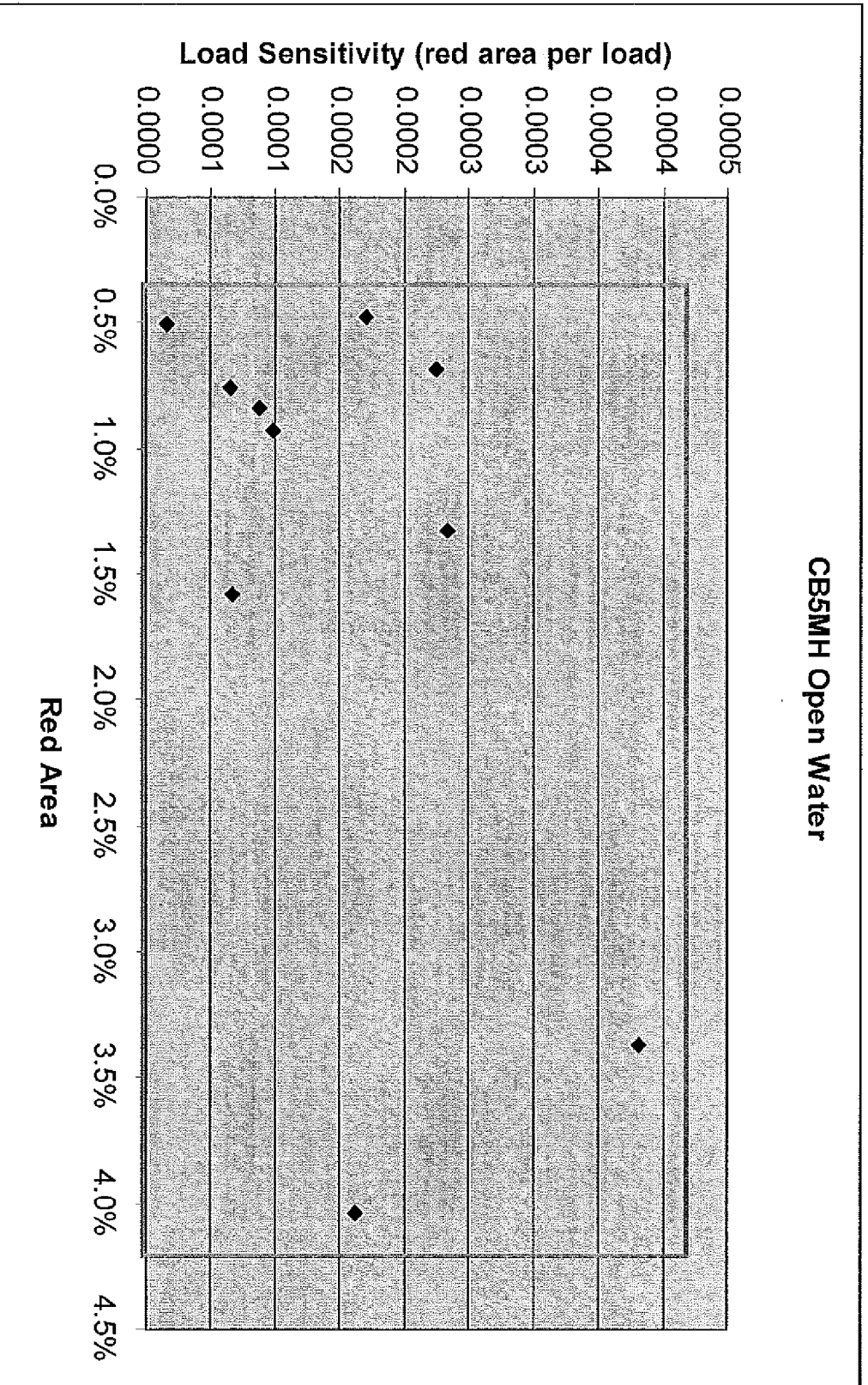
Sensitivity to Load Reductions

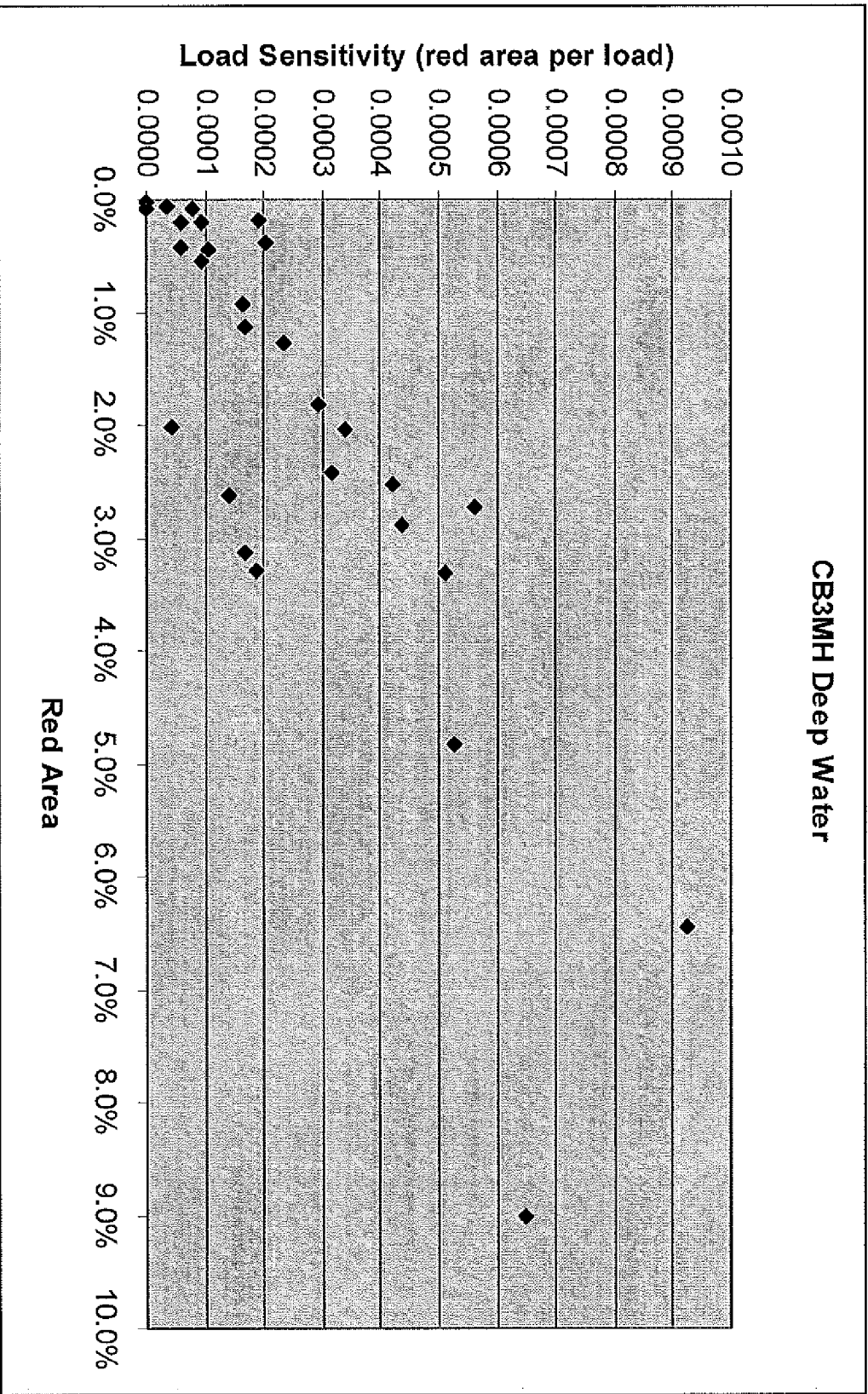
- Focused on segments driving the Bay TMDL
 - CB3MH, CB4MH, CB5MH for DW and DC and POTMH for DW
- Provided two open-water examples to show contrasts
 - CB2OH: drop in sensitivity at low non-attainment values
 - CB5MH: sensitivity to load reductions relatively constant throughout model simulated range



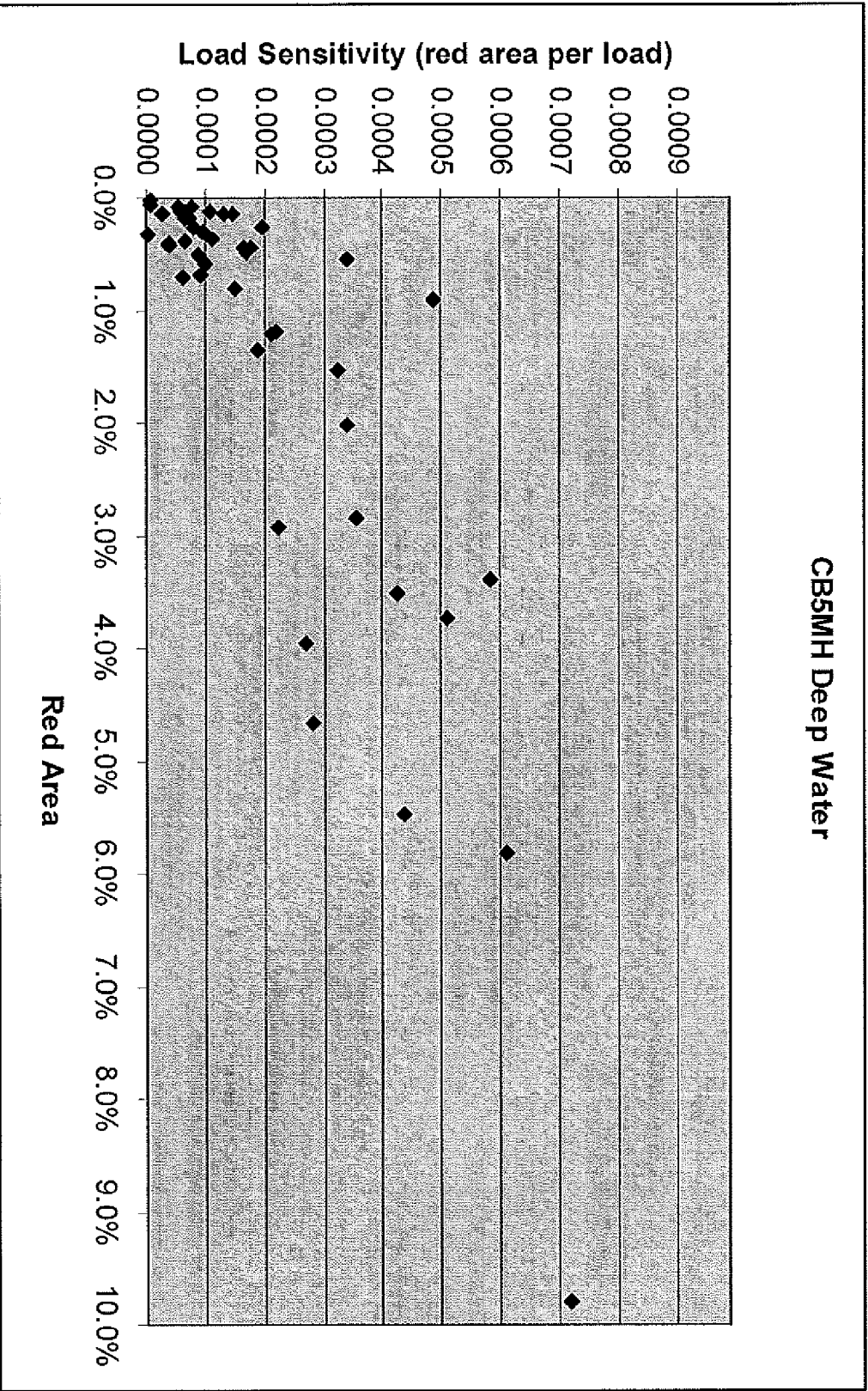


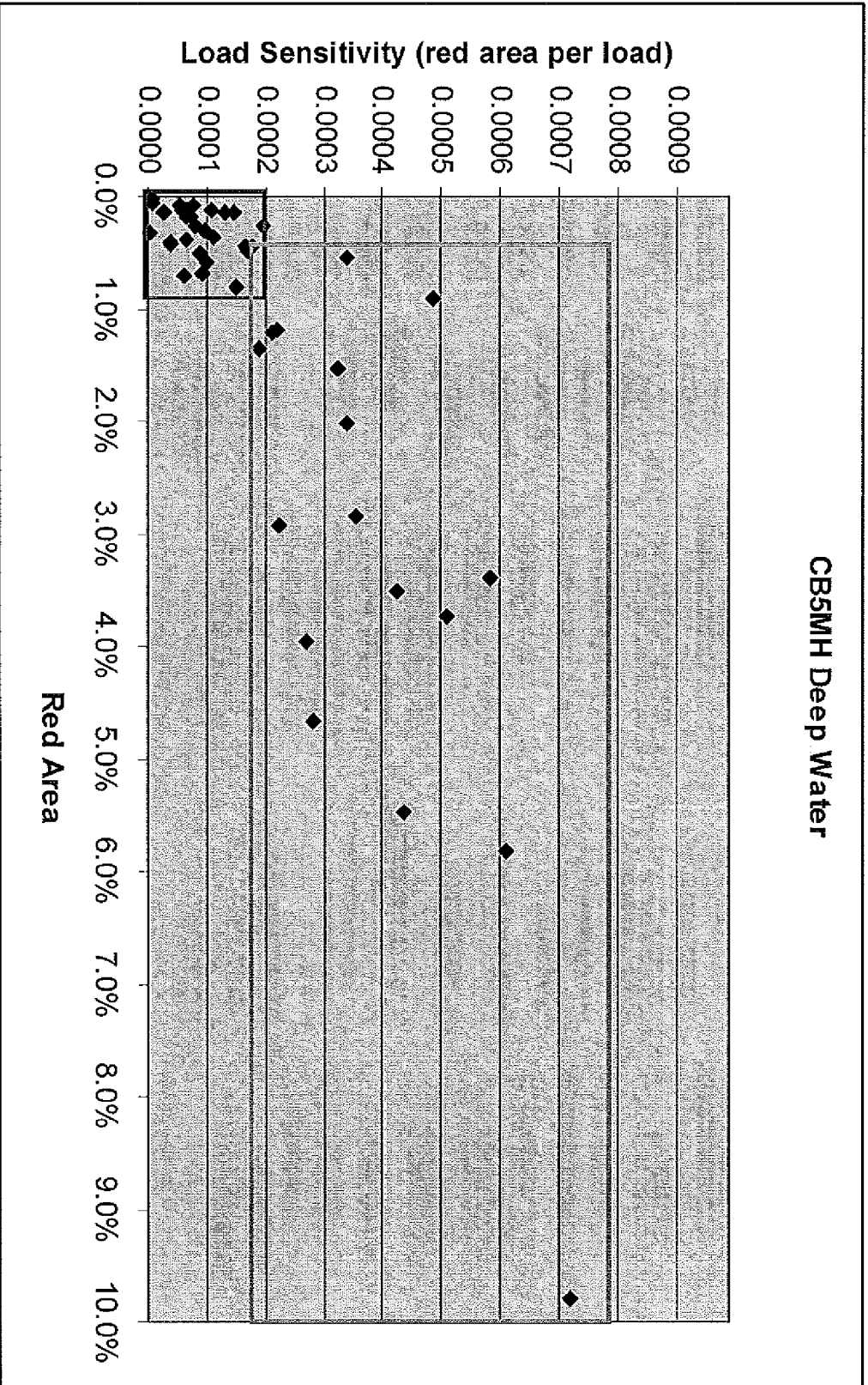


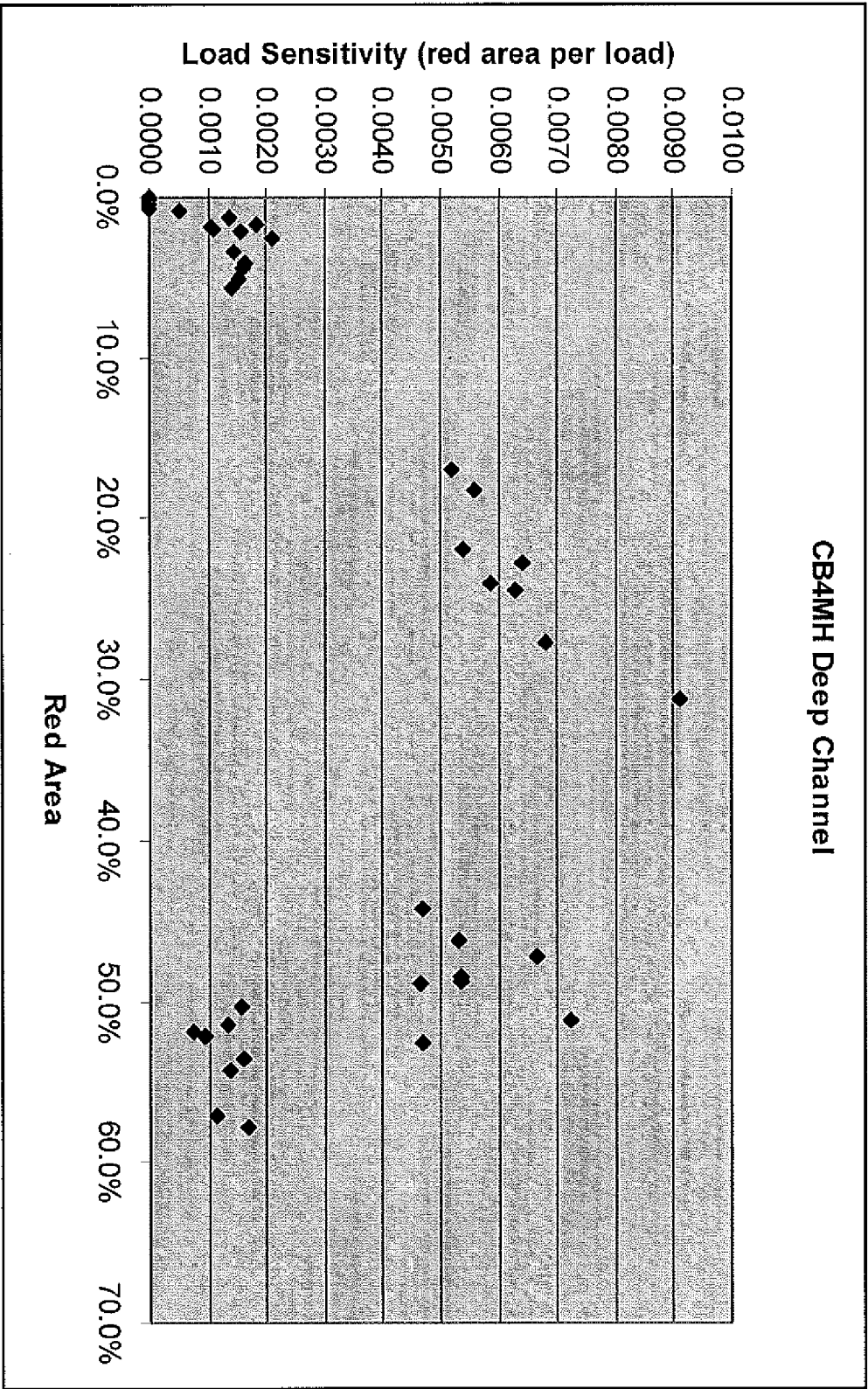


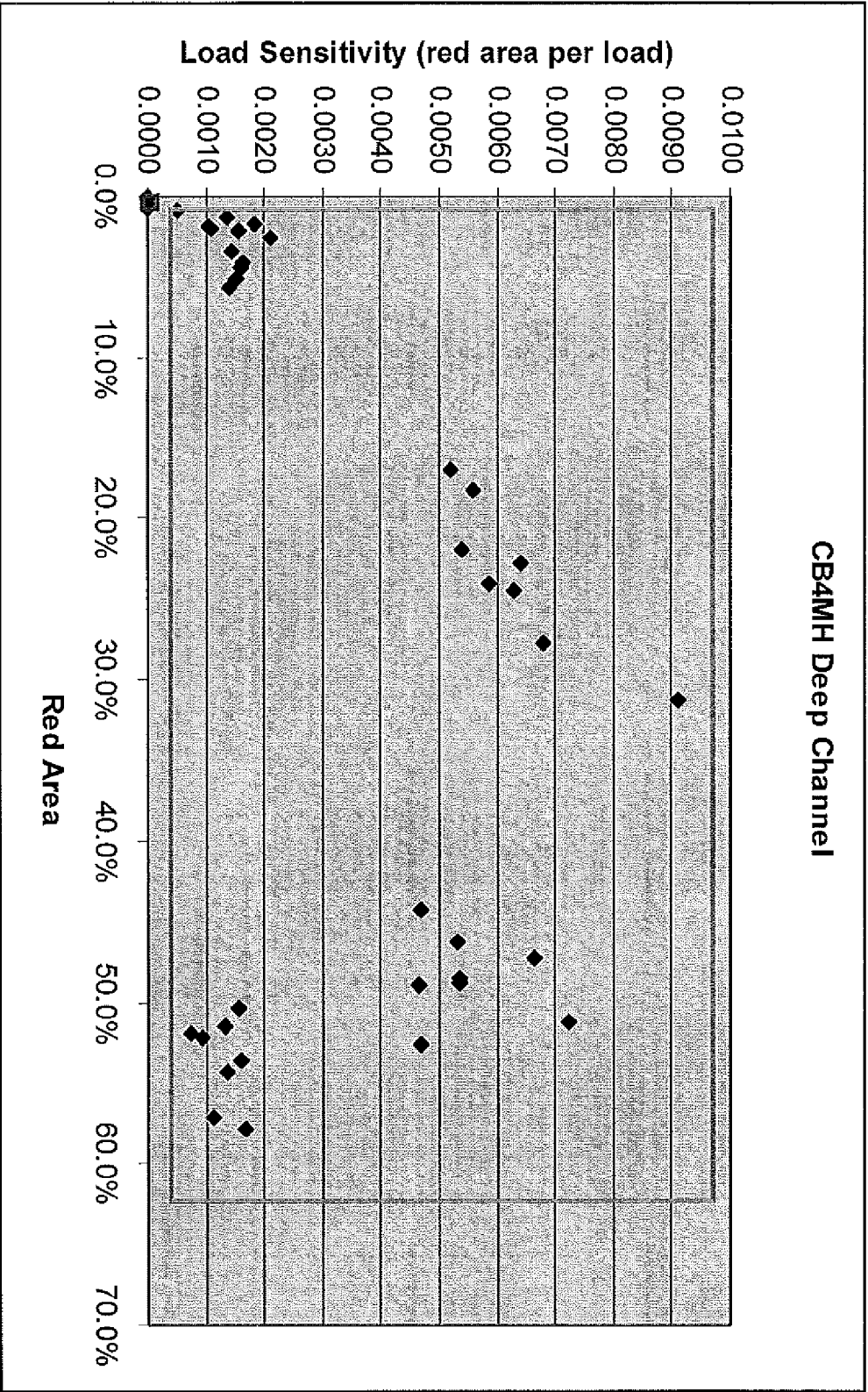


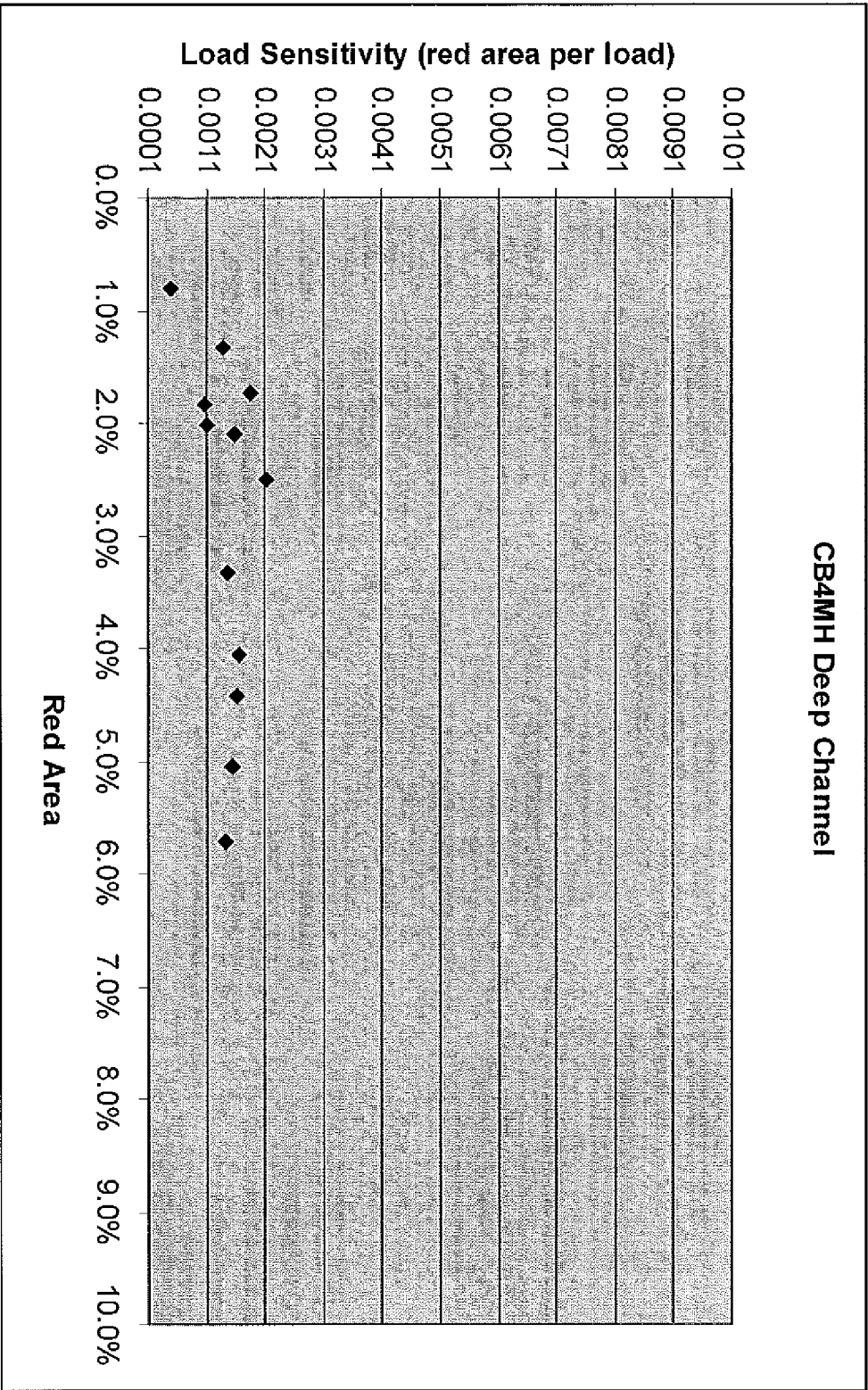


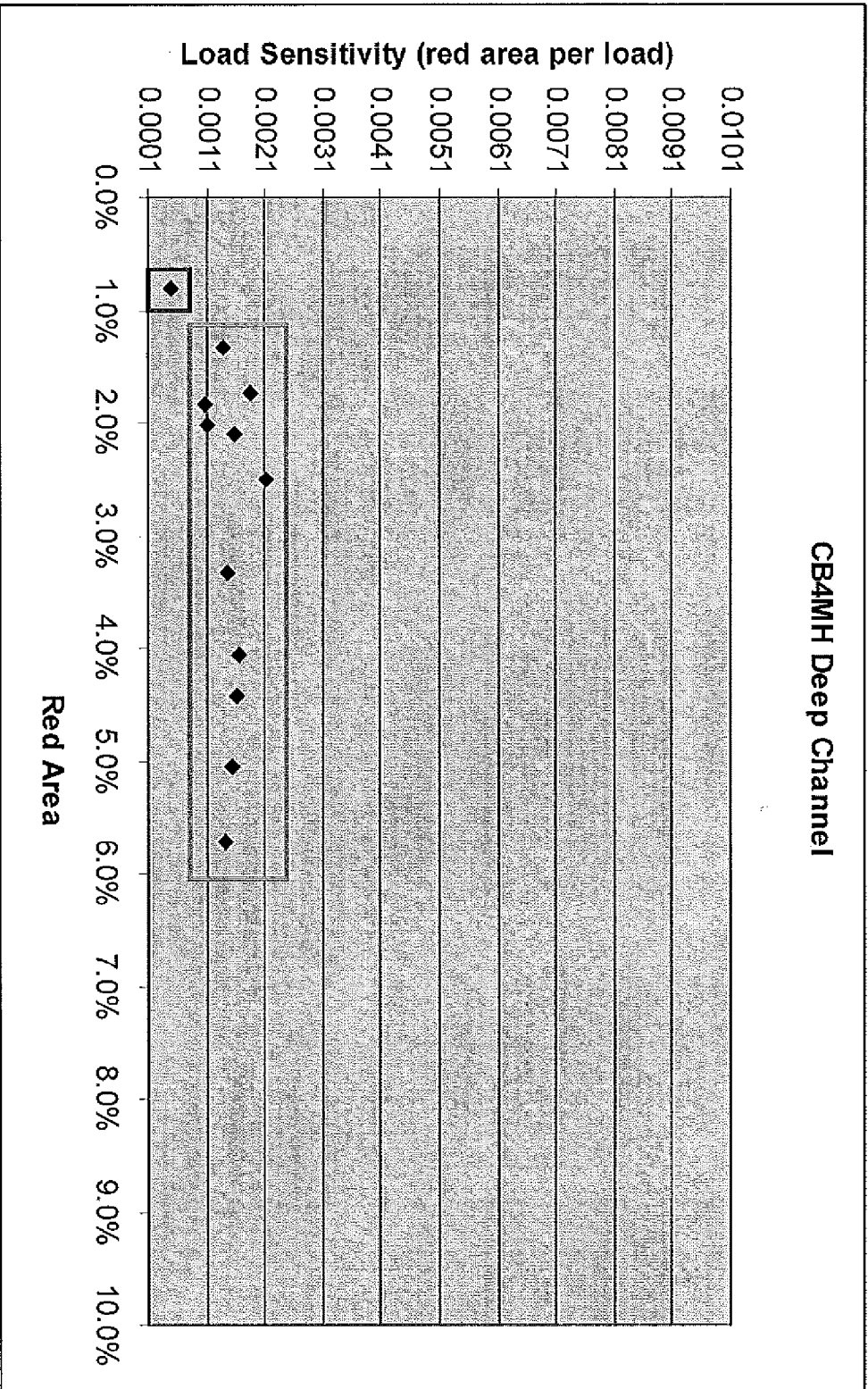












Chesapeake Bay Segment	Designated Use	Red Area with Low Sensitivity to Load Reductions (%)
CB3MH	Deep-water	0.2
CB4MH	Deep-water	0
CB5MH	Deep-water	1
POTMH	Deep-water	1
CB3MH	Deep-channel	1-1.5
CB4MH	Deep-channel	1
CB5MH	Deep-channel	N/A

Attainment at 1% Non-Attainment

- 21 designated use-segments with non-attainment values ranging from 0.0% to 1.5% across wide N load reductions
- Significant drop-off in sensitivity to load reductions documented at low non-attainment percentages
 - 1% being the consistent level at which sensitivity decreases across most of the designated use-segments driving the Bay TMDL

Attainment at 1% Non-Attainment

- The analysis system has been shown to be significantly less sensitive to the effects of further model simulated load reductions at and below the 1% non-attainment level
- These findings support documentation of attainment of designated use-segments with model simulated dissolved oxygen criteria non-attainment at 1%
 - No evidence documented for either analysis supporting a higher percentage (e.g., 2-3%)